CIE Review for Bering Sea and Aleutian Island flatfish, 18th-20th of April 2017, Alaskan Fisheries Science Center, Seattle, Washington

Sven Kupschus

Prepared for:

Center for Independent Experts (CIE)

Independent System for Peer Review

Centre for Environment, Fisheries and Aquaculture Science

Lowestoft Laboratory

Pakefield Road

Lowestoft

Suffolk NR33 0HT

United Kingdom

Telephone: +44 1502 524454

e-mail: sven.kupschus@cefas.co.uk

www.cefas.co.uk

Contents

1 EXECUTIVE SUMMARY 3

1.1 Activities 3

1.2 Assessment evaluations 3

2 BACKGROUND 5

3 REVIEW ACTIVITIES 5

4 FINDINGS 5

4.1 Bering Sea and Aleutian Islands Arrowtooth flounder 5

4.1.1 Evaluation of the ability of the stock assessment model for arrowtooth flounder, combined with the available data, to provide parameter estimates to assess the current status of arrowtooth flounder in the Bering Sea and Aleutian Islands. 5

4.1.2 Evaluation of the strengths and weaknesses in the stock assessment model for arrowtooth flounder. 6

4.1.3 Evaluation of the assumption that male natural mortality is higher than female in arrowtooth flounder. 9

4.1.4 Recommendations for further improvements to the assessment model. 9

4.2 Bering Sea and Aleutian Islands Kamchatka flounder 11

4.2.1 Evaluate stock assessment approach to model the Kamchatka flounder resource using three spatially distinct trawl surveys to provide reliable estimates of productivity, stock status, and statistical uncertainty for management advice. 11

4.2.2 Evaluate likelihood profile approach to estimate natural mortality rate (and suggest/provide alternatives?) 12

4.2.3 Evaluate how survey catchability estimates are derived based on assumptions about relative stock distributions. 12

4.3 Bering Sea and Aleutian Islands flathead sole 14

4.3.1 Evaluation of the ability of the stock assessment model for flathead sole, with the available data, to provide parameter estimates to assess the current status of flathead sole in the Bering Sea and Aleutian Islands 14

4.3.2 Evaluation of the strengths and weaknesses in the stock assessment model for Bering Sea/Aleutian Islands (BSAI) flathead sole. 15

4.3.3 Evaluation of alternatives to the current length-based survey selectivity curves used in the assessment. 17

4.3.4 Potential evaluation of an equivalent BSAI flathead sole assessment model in Stock Synthesis 17

APPENDIX 1: Bibliographic References provided for the review. Note Doyle et al. In Prep could not be made available to reviewers in time for the meeting. 19

APPENDIX 2: Statement of work for Sven Kupschus 20

APPENDIX 3: List of panel participants: 29

1 EXECUTIVE SUMMARY

1.1 Activities

The review of flathead sole, arrowtooth flounder and Kamchatka flounder in the Bering Sea was conducted from the 18-20th April 2017 at the Alaska Fisheries Science Center (AFSC) in Seattle. Presented were overviews of the various data collection programs and each of the three individual assessments. Reviewers had been provided with the 2016 assessments, with presentations of the latest developments provided at the review.

1.2 Assessment evaluations

All assessments employed statistical catch at age models developed in ADMB as separate templates. Flathead sole proposed transitioning to an SS3 implementation structurally consistent with the previously accepted model. All models used the available catch data (including length / age information) and three surveys (BS-slope, BS-shelf and Aleutian Islands (not included in the proposed assessment for Flathead sole). Historically the species have not been considered target species and catch information has not historically been separated by species. Arrowtooth and Kamchatka flounder have only been identified reliably in surveys since 1991.

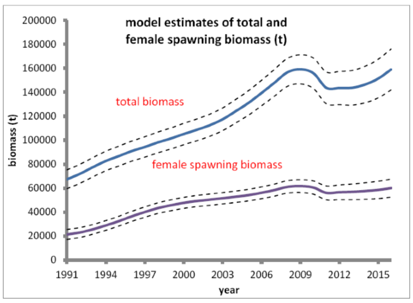
All three assessments suggest that the species are lightly exploited over the assessed period, and their SSB has increased several-fold over the period under low, occasionally annually spiking fishing mortalities. Length and age information has improved significantly over the period from historically quite poor data. Forward projecting models such as those employed here are somewhat susceptible to poor historic data, but VPA-based models are not suitable where historic age information is lacking. Survey data is pivotal to all three assessments, and in the historic period at least it is virtually replicated in the assessments in terms of the biomass trend because F is interpreted as being low. Internal consistency of cohort information in the raw data (both length and age) seems to be undervalued in the accepted assessments which suggest more stable recruitment than the raw data. Use of age-based rather than length-based selectivities alleviates the problem in some developmental models.

All assessments suggest that stocks are currently exploited sustainably, but were in an overexploited state in terms of their biomass (SSB < B40) at the beginning of the period. The assumption is that prior to the assessment F was much higher. Whether this is a credible interpretation is subjective. The main questions are, is the dramatic increase in flathead sole in particular likely? Would one not expect a more synchronous response from all the flatfish stocks? How can it match the chronology of the variation in catches of the main target species (exemplified by the fluctuations in halibut catches)? For arrowtooth flounder at least there is evidence that the historic fishery length distributions are similar to current ones, so that if SSB was historically low it would have been due to a period of poor recruitment rather than high Fs so that biomass reference points in relation to average recruitment may not be appropriate.

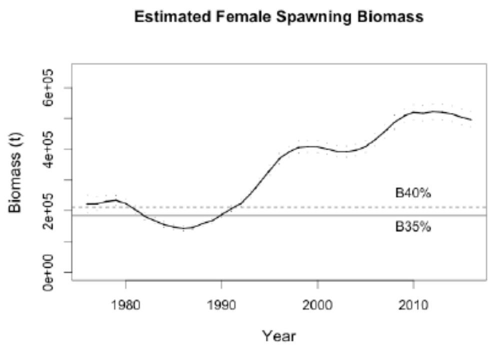
All three models seem to be towards the high end of model complexity given the data. Model convergence suggests that there is sufficient contrast in the recent data with the early period representing mostly a hind cast based on survey SSB. A better understanding of the appropriate choice of selectivity function is crucial to improving all of the assessments and should be the main focus of future developments. As well as modeling improvements, this will require continued or improved commitment to age information as the majority of the observed individuals have reached the growth asymptote where length provides little or no information on age, given the variability in length-at-age.

All three species are largely taken in multispecies fisheries predominantly as bycatch although there is some evidence of recent targeting of Kamchatka flounder for the Asian market. Management is by species specific quota, but on request of the industry it is possible to quota swap across species up to the ABC for a specific species, provided the sum of the OFLs across all species is not reached. The aim of the arrangement is to provide flexibility to industry to cope with unexpected fluctuations in species composition. The most influential management measure in effect for all three species is the restriction of 2 million tonnes of total extraction from the Bering sea. At present, more economically exploited species make up most of the catch. Under similar management, it is difficult to see how future exploitation of the flatfish complex would likely increase F beyond sustainable levels. More substantial data collection (see above) would provide a more certain assessment of the stocks, smaller confidence limits and allow exploitation closer to reference points (higher yield) at a given level of risk. At present other management strategies are more restrictive to the exploitation of these three flatfish species than the uncertainty around the assessment. Consequently, if the aim is to maximise the economic return from research / monitoring investment options other than more detailed flatfish assessments might be considered.

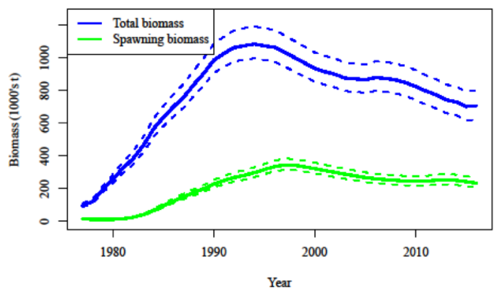
Flathead Sole



Kamchatka Flounder



Arrowtooth Flounder



2 BACKGROUND

My role as one of three reviewers participating in the review workshop was to read and evaluate the background reading materials, listen, comment and advise on the presentations, and give assistance in, and provide opinion on, ways to improve models. Following the meeting, I wrote a summary report of my findings and documented my recommendations. The following pages represent said report.

The process as a whole was an interesting and collaborative affair that focused mostly on improvement of models. Effectiveness of the review was occasionally hampered by being provided the 2016 assessments for pre-reading with some necessary information missing from those reports. Presentations given at the workshop were more elaborate, but given the number of assessments to review it was not possible to compare all diagnostics between different model settings and few additional options could be explored. If the hope was to maximise the opportunity to improve models, then more time to review the proposed experimental models either before or during the review would be useful as well as a clearer understanding of which model the TORs referred to. A more coherent presentation of model output within stock and across stocks would make it easier for reviewers to locate critical information when needed.

3 REVIEW ACTIVITIES

The 2017 BSAI flatfish review was held at the Alaska Fisheries Science Center (AFSC) in Seattle, Washington, from 18 to 20 April 2017. The bibliography consulted is listed in Appendix 1, and the Terms of Reference for the CIE panel in the SOW (Appendix 2: Annex 2).

A list of participants including panel members, SSC representation and observers are listed in Appendix 3. The meeting was open to the public, and was attended by one observer and intermittently by one industry representative via teleconference. In general, the questions asked of the reviewers were related to the assessment approaches and improvements to assessments and stock status determination rather than a specific evaluation of formal advice. For each stock, optional developments were presented to the panel while the 2016 accepted assessments were mostly for background reading. The conclusions in this report are largely based on this information, but some additional data / modelling requests were made by the panel which are not present in the original information.

4 FINDINGS

Findings are provided by species and in response to each term of reference (TOR).

4.1 Bering Sea and Aleutian Islands Arrowtooth flounder

4.1.1 Evaluation of the ability of the stock assessment model for arrowtooth flounder, combined with the available data, to provide parameter estimates to assess the current status of arrowtooth flounder in the Bering Sea and Aleutian Islands.

It was not clear to me which model we were being asked to assess. The last accepted model (2016), was provided prior to the review as report. Evaluating this model in detail was difficult due to the limited number of diagnostics included and some key errors in the evaluation of residual plots. The presentation at the meeting was mostly concerned with the development of an improved model. It did provide some information that was either directly from the 2016 model but more often was from a model that was similar to the former. The evaluation of the model given here is therefore more general to the modelling approach rather than a specific model choice. Generally, the model indicated a reasonable and tractable response to changes in the assumption / set up indicating that the likelihood gradients were sufficient to evaluate different options quantitatively. Examination of different development options suggested that the 2016 model was estimating more parameters than necessary. More parsimonious alternatives indicated little in the way of effects on the management metrics or diagnostic quality. Rather than incorrect *per se*, it merely means the model is unnecessarily vulnerable to misinterpretation of the data.

It is therefore concluded that the modelling approach is generally suitable in terms of providing estimates of current stock biomass and F. As for other flatfish stocks examined in this review uncertainty regarding these estimates for progressively earlier years becomes increasingly uncertain. Estimates of the dynamics suggest SSB has been increasing under comparatively low recruitment levels and low though constant Fs. Either Fs must have been higher in the period immediately preceding the data, or recruitment was even lower in that period. The implications of the alternate hypothesis regarding stock productivity (average long-term recruitment) and hence appropriate biomass reference points are significant.

Early catch data suggested that the length composition has changed little compared to more current catch data, and appear to have been ignored in favour of the earliest BS-shelf survey biomass data. This adds uncertainty regarding the early biomass trend not included in the assessment confidence limits, as well as raising concerns of whether the current reference points are suitable. It is therefore difficult to assess whether stock status estimates can be realistically considered reliable despite being quantifiable with apparent certainty within the assessment itself.

4.1.2 Evaluation of the strengths and weaknesses in the stock assessment model for arrowtooth flounder.

As indicated in the previous paragraph the main weakness of the assessment in terms of assessing stock status is in understanding the stock dynamics immediately preceding the assessment period. Data density in this period is comparatively low compared to more recent times, and is either uninformative or potentially conflicting between data sources. Options for redressing this uncertainty through the provision of additional data seems sparse, although anecdotal information could at least provide qualitative information (on reliability of early survey estimates and or early fishery length compositions) as to the most likely reason for the model interpretation.

Natural mortality rates are comparatively high (females=0.2, males=0.35) when considered in the light of estimates for other flatfish species in the area and worldwide. Ecosystem information suggests that there are few species that regularly prey on arrowtooth flounder which should imply that M could be lower than that used for other flatfish species in the Bering Sea. Greater certainty in M estimates would significantly improve the quality of SSB estimates in absolute terms; however, because fishing mortality has been low with little contrast over the period it is unlikely that the assessment would be able to provide more reasonable estimates of M. Alternate choices of M would merely rescale the assessment and likely provide little difference in the stock status metrics indicating that from a management perspective developing a better understanding of M should be considered a lower priority. More important to ascertain would be the likely cause for the sex specific Ms. The mortality offset between the species interacts quite strongly with the sex specific selectivity estimates particularly when selectivity is considered by length (one selectivity for both sexes) due to the differential growth in males and females. When considering age-based selectivities (one selectivity per sex) means a greater number of parameters are estimated but sexual differences in M will be less confounded with selectivities. This offers an opportunity to investigate the need for differential estimates of M to explain the observed sex ratio in survey catches biased towards females. Discussed further under point 4.1.3.

The strength of the assessment is its likelihood based approach that makes it possible to use all available information in the assessment, an important element given the sparsity of data overall, and the changes through time in the collection of different types of data. Generally, the aim is to avoid any arbitrary weighting of data; however, in this case it presents a problem for the survey data. The assessment uses three non-overlapping surveys. At least in terms of biomass estimates these surveys represent very different proportions of the population. The largest biomass is found in the BS-shelf survey which shows the most dramatic increase in biomass over time while the AI and BS-slope surveys have much flatter trajectories with occasional peaks. One would therefore expect that with an increasing weight given to the shelf survey, an assessment model should emphasize the increase in biomass more. With the majority of biomass and individuals found on the shelf giving extra weight to the BS-shelf survey is qualitatively appropriate, but quantitatively difficult to determine.

The last accepted arrowtooth flounder assessment considered it important to match the observed female favoured abundance ratio in the surveys. An analysis was presented at the review to demonstrate that the ratios were survey specific and indicated trends over time. The original assessment had some difficulties in matching these observations so an additional penalty function was introduced into the model, and sex specific mortalities were introduced in order to force the model to reflect these observations. I consider that the additional penalty should be superfluous if the remaining model structure reflects the stock dynamics appropriately. Similarly, sex specific mortalities should be considered an alternate hypothesis here and the focus should, at least initially, be to see if the dynamics could be replicated through differences in selectivity. Selectivity, either implemented as single size selectivity where by the skewed ratio could result from differential growth of the species, or by having sex specific selectivities at age. If this fails to resolve the discrepancy in the observed sex ratio differential mortality could still be implemented in a final model. Penalizing for divergence from the observed sex ratio on the other hand represents some arbitrary weighting factor which potentially could obscure other important stock dynamics.

Parameterisation of selectivities appeared to be more generous than necessary to model the stock dynamics. The difficulty in simplifying the model is that all selectivity parameters interact through their influence on abundance estimates. Therefore, the order in which to tackle these issues is important, plus changes must be considered iterative. Simple is better is the first rule, if selectivities for different sexes or fleets / surveys are not significantly different, a single parameterisation should be preferred. Dome-shaped selection requiring additional parameters should be considered carefully and is to be avoided unless at least one selectivity can confidently be determined to be asymptotic, as otherwise it may lead to large cryptic biomass estimates. Selection by length reflects the process of selection by the gear unless it is hypothesised that differences in selectivity at age are due spatial segregation in ontogeny. However, using length selectivities means that age information in this model is translated to length information through a transition matrix. When growth is inappropriately modelled (such as not accounting for changes in growth over time) or where length-at-age is highly variable, length data contains significantly less information on ages. In these cases, age transition matrices are problematic as they tend to make it difficult to identify strong or weak cohorts. The Figure 2 below taken from the assessment presentation illustrates the point. The deviation of the 2016 data from the fitted model is greater than the deviation between adjacent ages for ages greater than 10. Based on length data cohort tracking is unlikely to be successful.

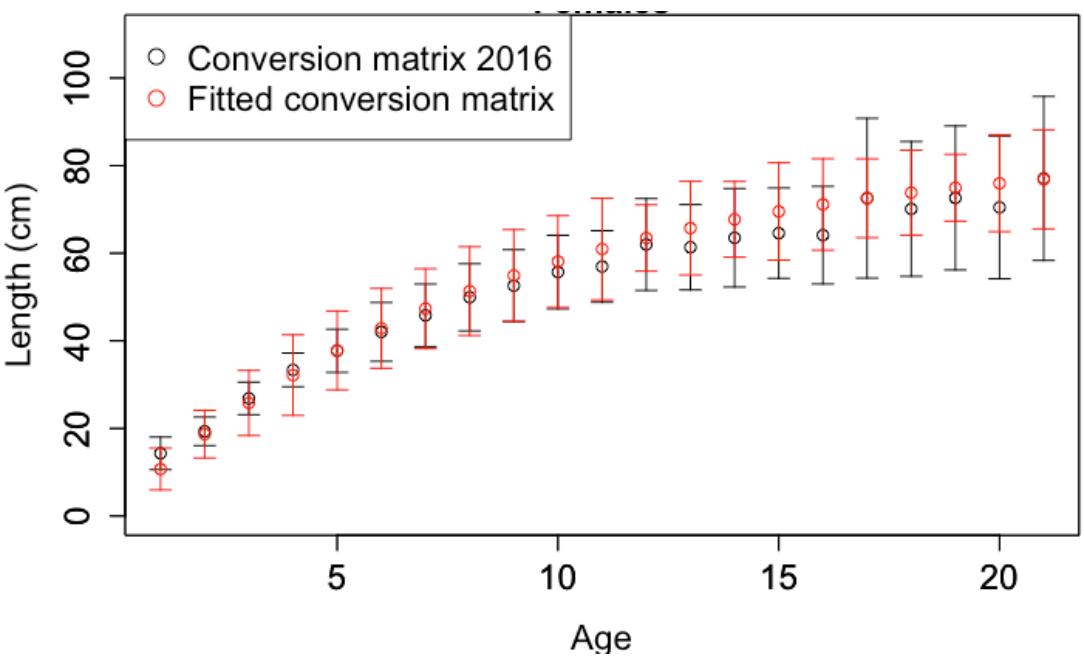


Figure 2: illustrating the difference between the observed and fitted length at age information. Important to note are the generally larger differences between the observed and fitted means at an age compared to the differences between fitted ages.

The panel was able to make some progress in simplifying the selectivities at the meeting, but there was insufficient time at the meeting to examine the effects of the model changes in detail. Next steps should be to look in more detail on the effect of removing the assumption of sex specific mortality and examine the effects of this on the selectivities, particularly the non-parametric fishery selectivity.

I was pleased to see that age information was appropriately treated in the model by accounting for the proportion of otoliths taken within a length group. Often age information is treated as random, despite the fact that it is collected length stratified. It was discussed that there were plans to move to random otolith sampling by observers in the future which would mean that the model would have to treat the old and new data differently. For many purposes collecting age information randomly is preferable, but for stock assessments tends to be more inefficient, since older individuals that are more difficult to distinguish based up on their length (particularly in the asymptotic part of the growth curve), tend to be rarer in the population. In contrast, length data on smaller more abundant individuals contains more information about age, because fewer cohorts overlap at a given length. Selecting ototliths randomly means one has to sample more small individuals to attain the same information on cohort strengths at the older ages. This can have budgetary or assessment quality consequences.

The currently accepted assessment uses a temperature anomaly adjustment on survey catchability. The model seems to be able to estimate this very precisely, although quantitatively it has little effect on the stock dynamics or management measures. Part of the reason for including this in the 2016 assessment is the wish to include ecosystem information in stock assessment models. To me this does not present ecosystem information in the sense that it does not affect the stock directly, only our ability to observe it as it is modelled as an effect on catchability. The effect is modelled as an annual average, whereas in reality it should be affecting different parts of the population separately. Lastly the fit is ‘unconvincingly good’, which suggest to me that despite its apparent statistical significance, it is an artefact of model over parameterisation and that the model uses relatively little information other than the biomass estimates of the survey so that it is able to accommodate any adjustment to survey indices, not just the temperature index. At present, this seems to have little effect on the management metrics, but the additional complexity constitutes a risk to future assessments. One way to assess this would be to bootstrap across the temperature index, examine the influence on stock dynamics and the significance of the effect.

4.1.3 Evaluation of the assumption that male natural mortality is higher than female in arrowtooth flounder.

Previous assessment assumed that a female biased sex ratio in catches is indicative of sex-specific natural mortality acting on the population. Although this is not without precedence in sexually dimorphic species with sex specific reproductive strategies, it is somewhat unusual for flatfish. Given that mortality estimates used in the existing assessment are high for flatfish anyway, the offset used seems unusually large. Wilderbuer and Turnock (2009) demonstrate sex-specific mortalities, but do not differentiate between fishing and natural mortality. If F is indeed as low as suggested by the assessment, then differential in the mortality rate of sexes is likely. Potentially, F may not be as low as estimated by the assessment in which case the assumption of higher male M may be pulling the model towards lower F estimates. I would first attempt to resolve the divergence through the suggested changes in selectivity (see 4.1.2) in order to gauge the effect. An investigation of spatial separation of the sexes might be useful to see if the surveys are representative of the population and to judge the appropriateness of the new selectivity curves. Second, the assumption of a 50% sex ratio at recruitment should be tested. It seems unlikely as a cause, because the bias in the sex ratio appears to increase with age but it could interact with selectivities. This recommendation is not to suggest that these alternative explanations of the sex ratio should ultimately be preferred, that will in the end depend on the diagnostics and the plausibility of the hypothesis examined in the different models. The primary aim is to examine the risks taken by making specific assumptions about the cause of the skewed sex ratio.

4.1.4 Recommendations for further improvements to the assessment model.

Several suggestions for improvement have already been made in the previous sections where they relate to specific TORs, but all suggestions should be taken into account equally when developing future assessments.

The major uncertainty regarding the arrowtooth assessments concerns the estimation of the status at the beginning of the time series. Although current stock status is comparatively robust to different model settings, conclusions about the historic status do influence the choice of appropriate reference points and management measures. The main reason for this uncertainty is the relative sparsity of information regarding the age structure at that time (implying a heavily exploited population at this time) based on low biomass values in the survey. However, the fishery length composition data early in the time series is very similar to the current length composition data suggesting a different initial age structure.

Anecdotally, there is a suggestion that historic Fs under different management regimes could have been higher, but generally the data has not been explored extensively. It seems unlikely that the data will be suitable for formal inclusion in assessments, but could qualitatively guide model development. The species is not considered a target species and therefore changes in the management of target species are likely to reflect changes in arrowtooth catch through changes in fleet distribution as much as through changes in abundance. Examination of the historic catches of target species with which arrowtooth flounder is caught should be undertaken to establish if historic Fs were likely to have been as high as suggested by the initial age structure estimated in this assessment, or if more weight should be given to the fisheries length distributions which suggest Fs were lower. As an example, international halibut catches do not imply a significant change in the catches and presumably F through changes in management (Figure 3).

Weighting of survey biomass is effectively done through fixing q's at 0.79, 0.09 and 0.13 for the shelf, slope and Aleutian island surveys based on the average biomass estimates.

Arrowtooth flounder and Kamchatka flounder exhibit similar distributions and exhibit a similarity in size and shape which would suggest that they are co-exploited irrespective of changes in the fishery through management. A common F trajectory for both species could be investigated by combining the species in a single assessment process that evaluates the species separately, but uses common F deviates and possibly selectivities to reduce the number of parameters required. This would also avoid having to make assumptions in the Kamchatka flounder assessment regarding the species ratio in historically combined catches (ignored in this assessment due to the much smaller proportion of the former species).

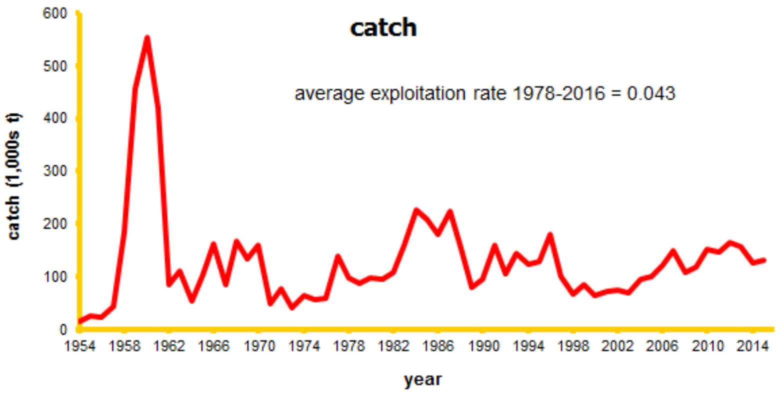


Figure 3: Halibut catches in the eastern Bering Sea as a rough proxy of historic fluctuations in fishing effort associated with the major management measures.

For most of the assessments in this review, the survey information, particularly the BS-shelf survey, is highly influential in driving the dynamics within the assessment. BS-shelf biomass indicates a steeper increase in SSB than the other surveys, so not only is it relevant to consider how to treat survey information in general, but also how to weight the three surveys appropriately. In the 2016 assessment survey catchability is set proportionate to the average annual biomass caught, and the sum of the catchabilities are constrained to 1.0.

Setting catchability to 1, implies that the survey index can be viewed as an absolute index of abundance. When presenting surveys at the meeting, survey scientists preferred to view survey estimates as a relative index of abundance. One of the reasons is that some avoidance is likely at all sizes, plus some areas, particularly the AI-survey are untrawlable due to either management restrictions or bottom obstructions. For these areas, abundance is estimated to be equivalent to the surveyed areas. Areas with management exclusion of survey and fisheries potentially have higher abundance, while untrawlable habitat could have lower abundance of soft sediment species such as flounder. It seems unlikely that the uncertainties will offset, despite their opposing effects. Some sensitivity examination should be performed to assess the sensitivity of the assessment to setting the sum of survey catchabilities to 1.

For the biomass estimate, weighting has been chosen to be proportional to the average annual survey biomass. This is appropriate where survey selectivities are identical. Weighting should be proportional to the fraction of the population covered, not the proportion caught. The former is of course unknown or at least dependent on the choice of the weighting within the assessment. As a first approximation, the choice of biomass based proportionality seems reasonable particularly given the similarity of the survey gears, but if selectivities are estimated to be different between surveys this assumption should be reviewed.

More complicated is the consideration of the lambda weighting of the survey length and age compositions. As with the biomass estimates weighting should be proportional to the fraction in the population in this case at length, or age. In the assessment, weighting is treated constant across all length and age distributions. As with the catchabilities data interpretation becomes complicated when selectivities differ. For example, if distributions are distinct in the survey areas, rare individuals may carry disproportionate weight, while abundant classes are likely to be under represented. There is little choice but to implement some arbitrary weighting, but as with the catchabilities it is reasonable to expect some assessment of the sensitivity to evaluate the uncertainty involved in making this choice.

4.2 Bering Sea and Aleutian Islands Kamchatka flounder

4.2.1 Evaluate stock assessment approach to model the Kamchatka flounder resource using three spatially distinct trawl surveys to provide reliable estimates of productivity, stock status, and statistical uncertainty for management advice.

The presented assessment for Kamchatka flounder presents a detailed examination of the information content of the available data for this stock. The modelling framework used is able to replicate a wide variety of population dynamic and sampling processes with the ability to statistically evaluate and contrast different options and data assumptions. In this sense, the model is state of the art. The available data in contrast is very limited, and potentially unable to support such complex evaluations. Specific reasons for this are:

a) The timeseries for which the Kamchatka flounder data can be reliably isolated from its congener arrowhead flounder is limited, especially in the catch data and a constant ratio assumption based on the 2007 data are used to hind cast catches by species to 1991, where the species is reliably estimated in survey catches. An analysis presented based on the % presence of each species in the survey hauls in the EBS survey indicates a high degree of correlation, but this could merely reflect the habitats sampled in the different years, and it indicated a small but persistent decline in the differences between the two species. Further evaluation of the species ratio in their abundance, rather than presence, would provide an important sensitivity check about this assumption.

b) The main reason for splitting the original species complex was given as concern over a developing target fishery due to market developments. The implemented model struggled to estimate survey fishery and survey selectivities without implementing some constraints. It does not seem sensible to assume that the selectivity in a historical by-catch fishery is likely to resemble those of a target fishery being persecuted largely in deeper areas known to be occupied by older larger individuals than the by-catch fishery. Unfortunately, the period under target fishery selectivities is insufficient to model this change in the fishery dynamics. With greater data density in the historic period / longer timeseries under by-catch conditions evaluating the risk of the assumption would have been possible, but with the current data this cannot be achieved.

c) Contrast in catches (aside of 2010) and survey estimated spawning biomass is generally small. This significantly complicates ranking the importance of different potential processes affecting the stock dynamics. The inability to effectively prioritise the dynamics is particularly problematic given the information content of the data which means only a limited number of parameters can be estimated before strong correlations between parameters becomes a problem.

4.2.2 Evaluate likelihood profile approach to estimate natural mortality rate (and suggest/provide alternatives?)

Profiling across different values of M is a common procedure to evaluate the risk to management of assumptions about M. The need for this approach frequently arises, because M is notoriously difficult to estimate when contrast in F is low and when M represents a large proportion of Z. Examining the likelihood profile to evaluate a potential choice of M is less common. The reason is that if this provides a conclusive answer as to which choice is optimal, then it should be possible for the model to estimate M appropriately within the minimization procedure. If the model fails to converge under these conditions, then it suggests likelihood gradient is insufficient to support such a decision.

As indicated under 4.2.1, the information content in the data is low (little contrast in F and low values of F) therefore it seems unlikely that the data supports a statistically significant discrimination of potential M values. I am not suggesting the choice of M is unreasonable, and in fact it is consistent with the estimates provided by other suggested methods based on max age or size. The risk in choosing the likelihood reasoning is that the current profile is conditional on the choice of other assumptions in the model. Because M is likely to be correlated with other parameter estimates, status estimates and management reference points would continually need to be revised.

Not examined in the assessment report was age dependent mortality traditionally based on growth parameters. Process wise, it is an intuitive alternative to the presented methods, but realistically in this dataset the effects on management metrics are going to be practically indistinguishable. The late recruitment means that the majority of the expected variation in M (at the youngest ages) will simply rescale recruitment estimates, while for older ages M will be correlated with selectivity parameters.

Although the evaluation of M is generally commendable, in this case it seems largely irrelevant, because there are other uncertainties in the assessment that are likely to have significantly more impact on the management metrics than the choice of M over a reasonable range. More importantly, given the current approach to management with the TAC always lower than the ABC due to the 2 million ton catch limit in the BSAI management plan mean TAC's are unlikely to change significantly even if the choice of M is likely to cause some variation in ABC estimates.

4.2.3 Evaluate how survey catchability estimates are derived based on assumptions about relative stock distributions.

a) Using relative biomass estimates from the survey to estimate q ignores the fact that the age composition differ. Survey biomass estimates are on a different scale than the catchability estimates that are based on log-abundance estimates. If the survey selectivities were more similar it might work as a first approximation. Slope survey mainly adults, EBS mainly juveniles, and AI a mixture. An abundance comparison, ideally at size, would be more useful. (See also discussion on the topic for arrowtooth flounder).

b) Acknowledging that there may be some data errors and resolve the issue by fixing q's for the BS-slope survey only fixes the problems in terms of the parameter estimation. The other survey biomass estimates are still incorrect and having fixed selectivities means other parameters are likely to suffer from bias as a consequence. It is a bias problem not a variance problem.

c) Slope survey parameters are more stable because the survey monitors mainly the oldest ages where cohort contrast is reduced. Therefore, the stability is inherent rather than a property of the survey *per se*. Considering the survey most reliable because of low variation seem risky to me. The BS-shelf survey is most independent of the fishery selectivity (predominantly younger ages), has significantly more length information over a greater number of years and lastly seems to provide reasonable estimates of recruitment, suggesting that abundance at age estimates are reliable making it the most information rich source to set selectivity parameters.

d) Catchability and natural mortality parameters are correlated. Fixing M based on profiling over q and then profiling q over the chosen value of M seems to be a circular argument. If indeed there is a distinct minimum in the likelihood, then the parameters should be estimable within the model. The two-step iterative approach is unlikely to reach such a minimum if it were to exist. However, management interest in the assessment appears to be SSB/biomass reference point. That relationship (for this stock) is relatively immune to the uncertainty caused by the correlation between M and q, so from a practical perspective, the two-step iterative approach seems a viable option.

Things to consider, investigate or explain:

Transforming length information into age information is facilitated through a probability of length-at-age transition matrix, based on one sex-specific growth function across all years. When length offers little information at age, i.e. the size-at-age overlaps considerably as it usually does in long-lived species, small variation in growth amongst cohorts, or high variability in length sampling tend to obscure cohort structure tending to dampen out cohort contrast. Even with only a modest amount of age data treating it multinomially should be preferred. Here age data appears to be restricted to a small number of years implying that it is not possible. In addition, the cohort signal based on length coming from the BS-shelf survey seems to be successful in identifying strong and weak year classes justifying the approach taken. However, significant damping of the cohort contrast cannot be excluded as a consequence. For status estimates this is likely to be of less concern, but forecasts will be more susceptible. Long-term management is likely to underestimate the threat to sustainability when the risk is shared amongst fewer cohorts than assumed in the model.

The model output suggested the biomass and spawning stock biomass estimates were at a minimum at the beginning of the time series, increasing rapidly over the first ten years. Neither a strong recruitment pulse nor any indication of a change in F were apparent from the assessment, so it is unclear what had disturbed the equilibrium conditions. Examination of the age structure predictions in 1991 indicated low population numbers at the older ages which explained the phenomenon in terms of the dynamics. What was less clear is what was driving the observed trend. The BS-shelf survey data provided no information on the age structure at the beginning of the timeseries, yet the estimated age structure in year 1 was interpreted by the model as being devoid of older ages. Subsequent rises in biomass result from the filling out of this age structure through below average cohorts at low Fs. It is unclear if the depauperate age structure is real or a model artefact. If it is real it is important to consider what caused the contracted age structure, despite the fact that it appears not to be influential on current biomass estimates (See also section 4.1.2 for options).

Estimated low Fs and lack of recruitment suggest that the stock could collapse to biomass levels well below current biomass reference point under minimal fishing pressure. If significantly higher historic Fs (possibly through poorly documented foreign fleet exploitation) had depleted the stock, then current biomass estimates are still likely to be appropriate, while low historic Fs would suggest that reference point should be reconsidered as this implies historically lower recruitment. If it is a model artefact, it is necessary to examine if the poor estimation has impacts on other parameter estimates (such as selectivity for example) that affect the current management metrics before current reference points can be evaluated.

4.3 Bering Sea and Aleutian Islands flathead sole

4.3.1 Evaluation of the ability of the stock assessment model for flathead sole, with the available data, to provide parameter estimates to assess the current status of flathead sole in the Bering Sea and Aleutian Islands

The flathead sole assessment provided as background reading was discussed only briefly in terms of it suitability to provide management advice. Its ability to provide estimates of stock status to assess the flathead sole is therefore largely based on the examination of proposed changes to the current model rather than evaluation of the models itself. A large effort had been made by the assessment scientist to move the model from the current *ad hoc* ADMB template file to an SS3 set up to simplify investigation and implementation of future models. There was insufficient time available at the review, nor did the review panel have the expertise to cross check the implementation of the template and the SS3 control file to ensure that both implementation were equivalent. Instead, it was only possible to do a gross evaluation based on the output parameters and a brief examination of the diagnostics. Structurally the model seemed to be doing very similar things, and trends in most of the output and diagnostics were very similar. One exception was the trend in F. In the original model this showed a steep decline from values of around 0.5 in the first years. This was replaced by a much smaller spike also levelling off, but at a lower absolute level than before. It was unclear if the plots produced were based on the same part of the selectivity curve or if the part used in the calculations provided a useful comparison.

It is unlikely that it will be possible to implement the original model in SS3 exactly. Almost certainly the way the models implement the internal data weighting will be different so that a reproduction of the assessment will not be possible. However, I do not think this particularly desirable given the wish to improve the assessment structurally. If still deemed necessary, further work is needed to understand the differences in F. Particularly, the trend in F early in the time series seem different (Figure 4). At first glance this is a small difference, but in terms of the interpretation, the differences seem more fundamental. The original model is interpreting the data as stock with an initially depauperate age structure requiring high levels of F to yield the observed catches at the time. The extremely steep increase in biomass is associated with the decline in F (i.e. rebuilding the age structure). The SS3 model instead seems to interpret the age composition in year 1 as more comprehensive (hence the lower initial Fs), but instead then estimates higher recruitment deviates early to replicate the biomass increase. From a management perspective, these results are very different. Comparison of the initial and terminal age structures of both models should indicate if this is the case. Also, stock status estimators should be compared between models.

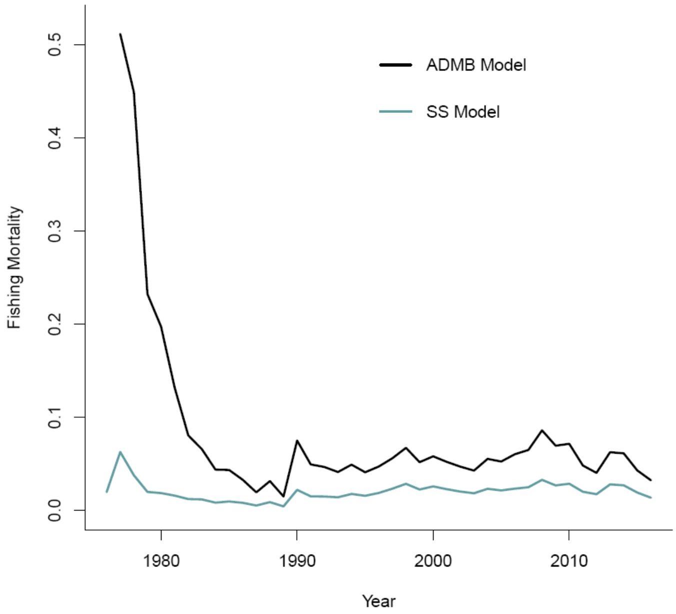


Figure 4: Difference in the F trajectories of the original and SS3 implementations. Some of the difference may be due to differences in the way F-deviates are output in the two models rather than real differences in the F estimates.

4.3.2 Evaluation of the strengths and weaknesses in the stock assessment model for Bering Sea/Aleutian Islands (BSAI) flathead sole.

The original model had several shortcomings, particularly with respect to the initial population dynamics. The steep decline in F trends seemed unlikely, given that the foreign cooperative fishery agreement had already been in place for some time prior to the assessment period. The exceptional cohort indicated in the survey was not well represented in the recruitment pattern. These issues were at least in part resolved by moving to an age based selectivity. The age length transition matrix seemed to complicate identification of cohorts at the older ages. Plots of mean length-at-age showed some temporal drift in the steepness of the VB curve (Figure 5) but the change seemed small, suggesting it was the variance in proportion of the overlap of cohorts in size. Switching to an age-based selectivity meant the age information was used directly where available avoiding the cohort smearing.

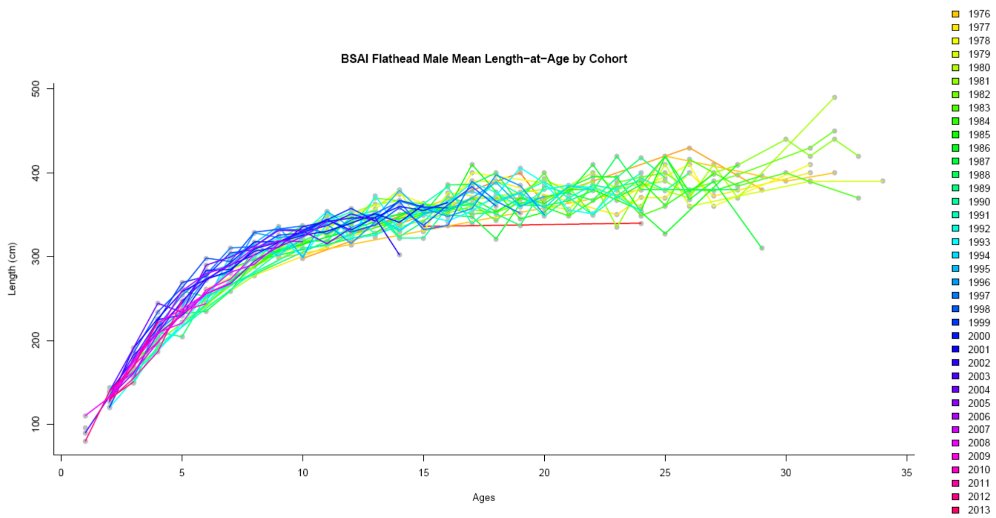


Figure 5 Mean size-at-age for individual cohorts indicating the similarity in variability within and between adjacent ages. Also indicated is a steeper initial growth phase for more recent cohorts.

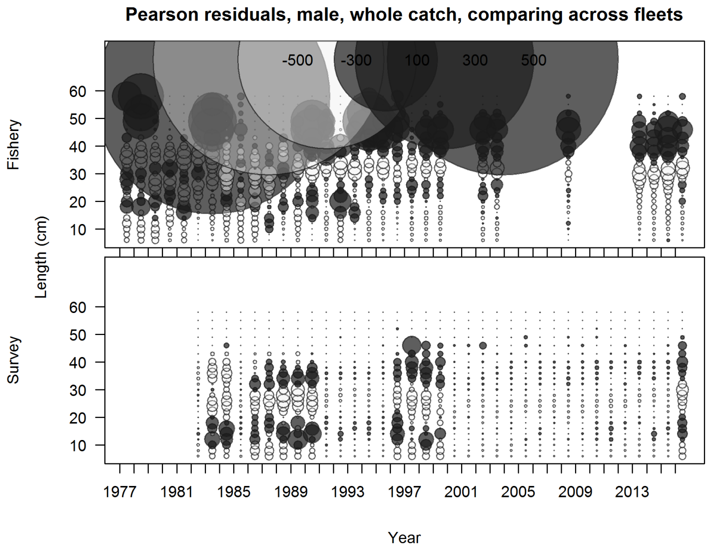


Figure 6 Male length residuals (as an example) indicating systematic distribution of positive and negative residuals indicating a poor fit to both the fishery and BS-shelf survey. Years with small residuals represent years where age information is available.

Despite these improvements, the model remained surprisingly unresponsive to what seemed like substantial further exploratory changes to model structure. There was some response to changes at the start of the time series. However, this was mainly due to a lack of information beyond biomass data to inform on F, rather than variation in the interpretation of the stock dynamics. It is unclear why the model is so stiff. Making changes to selectivities, introducing selectivity blocks, and even time variant selectivity had relatively little effects on residual patterns (Figure 6). It seems there is currently some conflict between the fishery age and age transformed length data. The very low F estimates may leave the model little room to reduce Fs further. If this were to coincide with an improving age structure, the model may be pushed up against its dynamic boundaries. It should be checked whether parameters are coming up against soft bounds. Age residual patterns suggested some rather complex selectivity curves would be required to randomise the residuals (Figure 6). There was insufficient time at the review to resolve these issues. The latest exploratory models are still somewhat unconvincing for the provision of advice.

The 2016 assessment model suffers from significant retrospective issues, overestimating SSB and underestimating F. Retrospective patterns are never good, but can be expected when as in this case historic information is comparatively sparse and model parsimony decreases as the final year's data is removed. Here though it seems the effect is not lessened as expected if the model was homing in on parameters, instead SSB-revisions appear to be of a consistent magnitude. This may be indicative of a more serious structural issue with the model and could be consistent with the idea that some parameters are pushing up against some bounds / penalty functions are not transitioning smoothly. With the complex retrospective patterns of the current and proposed model it is difficult to determine if choice of selectivity and retrospective patterns are linked. Sorting the selectivity problem should take precedence, but examining the reasons for the retrospective patterns may inform on a more appropriate choice of selectivities.

The AI-survey was not used in the assessment due to the small number of flathead individuals caught. Closer examination though did suggest that the survey, particularly in recent years had seen significant numbers of individuals. Moreover, these were predominantly small individuals which given the low catch rates in the other surveys, may present an important part of the recruiting population. Inclusion of the AI-survey should be investigated even if only to confirm that this does not change the perception of the stock. Unfortunately, including multiple surveys in a final assessment does pose the question how to appropriately weight these (also see sections on other stocks).

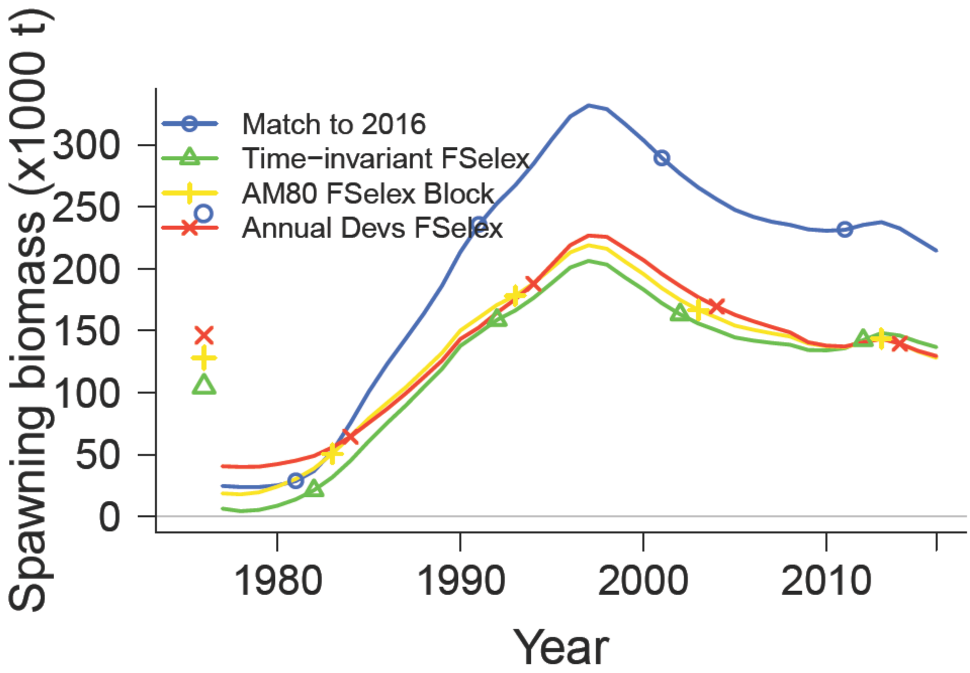


Figure 7 Significant changes to selectivity parameterisation indicated surprisingly little effect on the model estimates, in contrast moving from length-based to age-based selectivities (in blue) showed a more dramatic change.

Catches generally are very low compared to survey estimates of biomass suggesting there is little information on stock dynamics in the fisheries information. Appropriately then, the survey biomass is largely mirrored by the assessment reviewed here.

4.3.3 Evaluation of alternatives to the current length-based survey selectivity curves used in the assessment

The move to age based selectivities did improve the detection of cohorts, so this is almost certainly worth pursuing, but as indicated above it also increases the rigidity of the model and will require further work around understanding what selectivity functions might be appropriate (See also 4.3.1 and 4.1.2 on arrowtooth flounder in relation to the ability of models to detect cohort contrast based on length information).

4.3.4 Potential evaluation of an equivalent BSAI flathead sole assessment model in Stock Synthesis

Both the current and SS3 modeling frame works are principally aiming to achieve the same thing. Implementation of the modelling assumptions can vary slightly in SS3 from what is done currently, but the versions could easily be harmonized by changing either. The choice is a personal preference, with one being ultimately flexible (current) though more time consuming and more fraught with potential errors to implement new options. SS3 has many ready implemented options but they are still limited and at least historically documentation was poor in describing exactly how they interacted with other options. As indicated in section 1, it has so far not been possible to recreate the old assessment results in SS3 exactly. There is no definition of what differences should be thought of as material. The output metrics in F, SSB and recruitment are reasonable. Similar but much more substantial changes than the framework change indicated relatively minor responses in these metrics, so I am yet to be convinced that the models could not diverge in future given the same data. Scientifically, I do not consider it an issue as I do not find the old model particularly convincing. From a consistency and transparency perspective of the stock assessment process it is, however, undesirable.

APPENDIX 1: Bibliographic References provided for the review. Note Doyle *et al.* In Prep could not be made available to reviewers in time for the meeting.

Spies, I., Wilderbuer, T.K., Nichol, D.G. and Hoff, J,. Palsson, W., 2016. Arrowtooth flounder. Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions, pp.921-1012.

http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIatf.pdf

Doyle, M., Debenham, C., Barbeaux, S., Buckley, T., Spies, I., Pritle, J., Shotwell, K., Wilston, M., Cooper, D., Stockhausen, W., and Duffy-Anderson, J. In Prep. A full life history synthesis of Arrowtooth Flounder ecology in the Gulf of Alaska.

Wilderbuer, T. and Turnock, B. 2009. Sex-Specific Natural Mortality of Arrowtooth Flounder in Alaska: Implications of a Skewed Sex Ratio on Exploitation and Management, North American Journal of Fisheries Management, 29:2, 306-322, DOI: 10.1577/M07-152.1.

Wilderbuer, T., J. Ianelli, D. Nichol, and R. Lauth. 2016. Assessment of the Kamchatka flounder stock in the Bering Sea and Aleutian Islands. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK.
http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIkamchatka.pdf

NPFMC. 2017. BSAI Introduction. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK.
http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIintro.pdf

McGilliard, C.R., Nichol, D. and Palsson, W. 2016. 9. Assessment of the Flathead Sole-Bering flounder Stock in the Bering Sea/Aleutian Islands Regions. In Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. pp. 1229-1318. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIflathead.pdf

APPENDIX 2: Statement of work for Sven Kupschus

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Fisheries Stock Assessments for Arrowtooth Flounder, Flathead Sole and Kamchatka Flounder

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available. NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Alaska Fisheries Science Center's (AFSC) Resource Ecology and Fisheries Management Division (REFM) requests an independent review of the integrated stock assessments that have been developed for three Bering Sea flatfish species; arrowtooth flounder, flathead sole and Kamchatka flounder. The fishery for these species is managed by the North Pacific Fisheries Management Council. The sum of the Allowable Biological Catches (ABCs) for these three species is 142,529 t in 2017, with catch levels annually set lower than the ABC due to a 2.0 million t harvest cap for all species and constraints due to Pacific halibut bycatch limits and markets. The catch limits are established using Automatic Differentiation (AD) Model software that uses survey abundance data and survey and fishery age and length composition data with a harvest control rule to model the status and productivity of these stocks and set quotas. Having these assessments vetted by an independent expert review panel is a valuable part of the AFSC's review process. The Terms of Reference (TORs) of the peer review and the tentative agenda of the meeting are below.

Requirements for CIE Reviewers

NMFS requires three CIE reviewers to conduct an impartial and independent peer review in accordance with the SOW, OMB Guidelines, and the TORs below. The reviewers shall have working knowledge and recent experience in the application of fisheries stock assessment processes and results, including population dynamics, separable age-structured models, harvest strategies, survey methodology, and the AD Model Builder programming language. Experience with the Stock Synthesis Assessment Model would also be helpful. They should also have experience conducting stock assessments for fisheries management.

Statement of Tasks

- Review the following background materials and reports prior to the review meeting:

Spies, I., Wilderbuer, T.K., Nichol, D.G. and Hoff, J,. Palsson, W., 2016. Arrowtooth flounder. *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions*, pp.921-1012.

http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIatf.pdf

Doyle, M., Debenham, C., Barbeaux, S., Buckley, T., Spies, I., Pritle, J., Shotwell, K., Wilston, M., Cooper, D., Stockhausen, W., and Duffy-Anderson, J. In Prep. A full life history synthesis of Arrowtooth Flounder ecology in the Gulf of Alaska.

Wilderbuer, T. and Turnock, B. 2009. Sex-Specific Natural Mortality of Arrowtooth Flounder in Alaska: Implications of a Skewed Sex Ratio on Exploitation and Management, North American Journal of Fisheries Management, 29:2, 306-322, DOI: 10.1577/M07-152.1.

Wilderbuer, T., J. Ianelli, D. Nichol, and R. Lauth. 2016. Assessment of the Kamchatka flounder stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK.
http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIkamchatka.pdf

NPFMC. 2017. BSAI Introduction. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK.
http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIintro.pdf

McGilliard, C.R., Nichol, D. and Palsson, W. 2016. 9. Assessment of the Flathead Sole-Bering flounder Stock in the Bering Sea/Aleutian Islands Regions. In Stock Assessment and Fishery

Evaluation Report for the Groundfish Resources of the Bering Sea and Aleutian Islands. pp. 1229-1318. North Pacific Fishery Management Council, P.O. Box 103136, Anchorage, AK 99510. http://www.afsc.noaa.gov/REFM/Docs/2016/BSAIflathead.pdf

- Attend and participate in the panel review meeting
    - The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers
- After the review meeting, reviewers shall conduct an independent peer review in accordance with the requirements specified in this SOW, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus
- Each reviewer may assist the Chair of the meeting with contributions to the summary report, if required by the TORs
- Deliver their reports to the Government according to the specified milestone dates

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and at the Alaska Fisheries Science Center, Seattle, Washington.

Period of Performance

The period of performance shall be from the time of award through June 12, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
No later than April 4, 2017	Contractor provides the pre-review documents to the reviewers
April 18-20, 2017	Panel review meeting
May 8, 2017	Contractor receives draft reports
May 30, 2017	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each TOR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $10,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact:

Tom Wilderbuer

Tom.Wilderbuer@noaa.gov

National Marine Fisheries Service,

7600 Sand Point Way, NE, Bldg. 4,

Seattle, WA 98115-6349

Phone: (206) 526-4224

Final Agenda:

CIE Flatfish assessment review

NMFS Alaska Fisheries Science Center

7600 Sand Point Way NE, Building 4

Seattle, Washington

Agenda *FINAL VERSION* April 18-20, 2017

Tuesday April 18th

Room 2111, Building 3

9:00	Welcome and Introductions, adopt agenda	
9:15	Overview (species, biology, surveys, fishery, catch levels, ABCs, TACs, bycatch)	**Tom**
9:30	Bering Sea trawl shelf survey	**Bob Lauth**
10:00	Aleutian Islands trawl survey	**Ned Laman**
10:30	Bering Sea slope trawl survey	**Jerry Hoff**
11:00	Coffee break	
11:15	Observer Program	**Marlon Concepcion**
11:40	Age Determination	**Delsa Anderl**
12:00	Lunch	
	Room change to Traynor room in Building 4	
1:00	Effect of multiple management actions on flatfish fisheries	**Alan Haynie**
1:30	Bering Sea Kamchatka flounder	**Tom and Jim Ianelli**

Wednesday April 19th Room 2011 Building 4

9:00	BSAI flathead sole	**Carey McGilliard**
11:00	Coffee break	
11:20	BSAI flathead sole (continued)	**Carey McGilliard**
12:30	Lunch	
1:30	Bering Sea Arrowtooth flounder	**Ingrid Spies**

Thursday April 20th   Room 2011 Building 4

9:00	Bering Sea flatfish discussion
11:00	Coffee break
11:20	CIE panel discussion (assessment authors will be available)
12:30	Lunch
1:30	CIE panel discussion (assessment authors will be available)

Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review

   Appendix 2: A copy of this Statement of Work

   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Terms of Reference for the Peer Review

Bering Sea and Aleutian Islands Arrowtooth flounder

1. Evaluation of the ability of the stock assessment model for arrowtooth flounder, combined with the available data, to provide parameter estimates to assess the current status of arrowtooth flounder in the Bering Sea and Aleutian Islands.
2. Evaluation of the strengths and weaknesses in the stock assessment model for arrowtooth flounder.
3. Evaluation of the assumption that male natural mortality is higher than female in arrowtooth flounder.
4. Recommendations for further improvements to the assessment model.

Bering Sea and Aleutian Islands Kamchatka flounder

1. Evaluate stock assessment approach to model the Kamchatka flounder resource using three spatially distinct trawl surveys to provide reliable estimates of productivity, stock status, and statistical uncertainty for management advice.
2. Evaluate likelihood profile approach to estimate natural mortality rate (and suggest/provide alternatives?)
3. Evaluate how survey catchability estimates are derived based on assumptions about relative stock distributions.

Bering Sea and Aleutian Islands flathead sole

1. Evaluation of the ability of the stock assessment model for flathead sole, with the available data, to provide parameter estimates to assess the current status of flathead sole in the Bering Sea and Aleutian Islands
2. Evaluation of the strengths and weaknesses in the stock assessment model for Bering Sea/Aleutian Islands (BSAI) flathead sole
3. Evaluation of alternatives to the current length-based survey selectivity curves used in the assessment
4. Potential evaluation of an equivalent BSAI flathead sole assessment model in Stock Synthesis

APPENDIX 3: List of panel participants:

Anne Hollowed	AFSC Status of stocks
Carey McGilliard	AFSC Status of stocks
Ingrid Spies	AFSC Status of stocks
Meaghan Bryan	AFSC Status of stocks
Tom Wilderbuer	AFSC Status of stocks
Sandra Lowe	AFSC Status of stocks
Jim Ianelli	AFSC Status of stocks
Alan Haynie	AFSC Economics program
Jerry Hoff	AFSC Bering Sea groundfish survey
Bob Lauth	AFSC Bering Sea groundfish survey
Dan Nichol	AFSC Bering Sea survey program
Ned Laman	AFSC Aleutian Islands groundfish survey
Beth Matta	AFSC Age and growth program
Delsa Anderl	AFSC Age and growth
Marlon Concepcion	AFSC Observer program
Todd Loomis	Industry (Ocean Peace)